SOME OPTIMIZATION PROBLEMS IN MULTIVARIATE BAYESIAN SAMPLING

THESIS SUBMITTED
FOR THE AWARD OF THE DEGREE
Doctor of Philosophy
IN
Statistics

BY

ZAHIRUL ISLAM

DEPARTMENT OF STATISTICS
ALIGARH MUSLIM UNIVERSITY
A L I G A R H

1 9 8 1

C E R T I F I C A T E

I certify that the material contained in this thesis, submitted by Mr. Zahirul Islam, for the award of the degree of Doctor of Philosophy in Statistics is original.

The work has been done under my supervision. In my opinion this work is sufficient for consideration for the award of Ph.D. degree to the candidate.

Dated- 30.3.1981

(Dr. Sanaullah Khan)
Head,
Department of Statistics
Aligarh Muslim University
Aligarh, India.

P R E F A C E

This thesis entitled " Some optimization problems in Bayesian multivariate Sampling " is submitted to Aligarh Muslim University, Aligarh to supplicate the degree of Doctor of Philosophy in Statistics. It embodies the research work carried out by me in the Department of Statistics, Aligarh Muslim University, Aligarh.

The theory of Sampling is concerned with the development of most precise (or most economic) procedures of Sample selection and with the construction of good estimates of certain parameters of the population under study. One of these procedures is to divide the population into various strata.

In this thesis we are concerned with the allocation problem in multivariate surveys. The strategy used is to obtain minimum posterior variance of the mean for each of the p characters, so that the cost of the survey does not exceed the total available budget.

The problems in the various situations are shown to be reduced to that of convex programming with multiple objective functions.

Chapter I of the thesis gives the historical background and the concepts for the work presented in the sequal. The discussion helps to understand how the Bayesian techniques are useful for problems of optimum allocation in sampling.

In Chapter II a procedure has been developed for a programming problem with multiple objectives in which the objective functions are convex and the constraints are linear. A method for further improvements in the solution is also indicated.

In Chapter III the optimum allocation has been obtained in single phase sampling, when the population is normal and the prior distribution is also normal. There are more than one characters under study. Bayesian posterior analysis is employed.

In Chapter IV and V we treat the problem of optimum allocation, for multiple characters, for the second phase in a two phase sampling procedure. The population from which the samples are drawn is normally distributed. The preposterior and posterior analysis is employed and the posterior variances for all the characters are minimized subject to the cost restriction. In Chapter IV the prior distribution is non-informative for all the characters while in Chapter V, the prior distribution is conjugate.

I am extremely thankful to Dr. S. U. Khan, for his continuous supervision during this work. Thanks are also due to Mr. Nafees Alvi for his patient help in typing and Mr. Kishan Babu Kamal for his assitance in cyclostyle work of the thesis.

Zahirul Islam

Aligarh (Zahirul Islam)

Dated- 29.3.1981

C O N T E N T S

PAGE

Chapter I

Background and Preliminaries

1.1. Optimization and Approximation

(i) General Approximation Problem

Every approximation problem can be regarded as an optimization problem. Although, all theoritical problems of approximation theory can not be answered by means of optimization, the questions of practical importance, such as calculation of minimal deviation and the solution of approximation problem may be successfully obtained by using various methods of optimization. May approximation problems arising in applied mathematics and other fields have been shown for a possible treatment in a unified way within the general frame-work of optimization theory by Collatz, L. and W.Welleihing (1975), Krabs, W.(1979), Zukhovitskey, S.I.and L.I.Avdeyera (1966).

A problem of approximation generally has the following form : Given a continuous, real valued function, $f(x)$, defined on a given point set S, a class of real valued functions, $V = \{ g(x, a_1, \ldots, a_n) \}$, also continuous in x on S and dependent on the real parameters, $a_1, \ldots, a_n$, such that $f \notin V$ and a metric, $\rho(f,g)$, further distance between two functions, f and g, the problem is to find a function $g \in V$ for which the distance $\rho(f,g)$ is minimal. An approximation problem thus may be considered as an optimization problem with objective function $\rho(f,g)$.

(ii) Convex Chebyshev approximation

Assume that we are given a system of p convex smooth functions $f_i(x)$, $i = 1,2,\ldots,p$ and a region S defined by the inequalities $h_j(x) \leq o$, $j = 1,\ldots, q$ where $h_1(x),\ldots h_q(x)$ are also convex smooth functions. The corresponding convex Chebyshev approximation problem consists in finding a point $x^* \in S$ for which $\max_i f_i(x^*) = \min_{x \in S} \max_i f_i(x)$.

As the function $\max_i f_i(x)$ is convex, the convex Chebyshev approximation problem is also a convex programming problem.

(iii) Points closest to a system of hyperplanes

In n-dimensional Euclidean space, consider a system of m hyperplanes $h_i : a_1^{(i)}x_1 + \ldots + a_n^{(i)} x_n + a^{(i)} = o, i=1,2,\ldots$.

The quantity

$\rho(x,h_i) = \left| \sum_{j=1}^{n} a_j^{(i)} x_j' + a^{(i)} \right|$ is the distance from the point $x' = (x_1', \ldots, x_n')$ to the plane h_i, where $|\quad|$ represents the Euclidean norm. The problem is to find the point with least distance from the system of hyperplanes h_i, i.e. a point x^*, for which

$$\max_{1 \le i \le m} \rho(x^*, h_i) = \inf_{x} \max_{1 \le i \le m} \rho(x, h_i)$$

Again, as the function $f(x) = \max_{i} \quad (x,h_i)$ is convex, the above problem is also a convex programming problem. The above equivalence has been exploited in Chapter II for developing an algorithm for a convex programming problem with multiple objective functions.

1.2. Optimization with Multiple Objectives.

When we find an optimal solution to a Mathematical programming problem we actually suboptimize the problem of a larger system of which the problem solved was a component for acheiving the goal of the system as a whole the formulation in the form of programming with multiple objective is considered Benayonn, R and J. Tergny (1969) have given the following formulation in the linear case.

To find a vector $\underline{x}$ in the feasible region $D \in R^n$ defined by the solution of the linear system :

$$\underset{m \times n}{A}\, \underline{x} \leq \underline{b} \quad , \quad \underline{x} = (x_1, \ldots, x_n)$$

$$x_i \geq o \quad , \quad 1 \leq i \leq n$$

where the chosen vector gives '' Satisfaction '' to the p numerical characters denoted by c^1 , $c^2, \ldots, c^p$.

The value taken by a solution $\underline{x}$ corresponding to the character c^j is given by

$$y^j(\underline{x}) = \sum_{i=1}^{n} c_{ij}\, x_i, \quad j = 1, 2, \ldots, p.$$

It is assumed that more $y^j(\underline{x})$ is larger, more the solution $\underline{x}$ is satisfactory for the character c^j.

In the presence of some apriori knowledge regarding the relative importance of the various characters, the following two methods of recasting the problem is used.

Case 1. If the relative importance of the characters may be expressed by means of assigning weights, say, π^j for the j^{th} character, then a unique objective function can be defined as

$$F(\underline{x}) = \sum_{j=1}^{p} \pi^j y^j(x) = \sum_{i=1}^{n} d_i x_i ,$$

where $\quad d_i = \sum_j \pi^j c_{ij}, \; i = 1,2,\ldots, n.$

In this case we are led to an ordinary linear programming problem consisting of a unique objective function.

Case 2. If the apriori knowledge is available only in terms of an hierachy of the various characters, the characters are first arranged in order of priority. For example, in case of two characters let $y^1(\underline{x})$ have priority over $y^2(\underline{x})$. First we solve the problem

$$\max_{\underline{x} \in D} \; y^1(\underline{x})$$

Now let H represent the half space of solutions of the inequality

$$y^1(\underline{x}) \geq f' - \epsilon'$$

where $f' = \max\limits_{\underline{x} \in D} y^1(\underline{x})$ and ϵ' is a non-negative number representing the maximum allowable decrease in ϵ'. The required solution will then be the solution of the problem

$$\max_{x \in D \cap H} \; y^2(x).$$

The idea of a satisfactory solution is not very clear in the more general case where either the relation between the various characters, being of very diverse nature, is too complex or otherwise the available information is incomplete. A different method to cope with such situation is proposed with the name of POP and described in Benayonn, R. and J. Tergny (1969). The further renification of this method has been presented (again for the linear case) by the name STEP method in Benayonn, R, J. Montgolfier and Tergny, J (1971). The idea of STEP method has been exploited for developing a procedure for convex programming with multiple objectives in Chapter II.

1.3. Branch and Bound Methods

The branch and bound principle was first discovered by Land, A.H. and A.G.Doig (1960) in the context for solving integer linear programming problems. Later Dakin, R (1965) proposed a simple, yet interesting, variation of Land and Doig algorithm.

The basic idea of the branch and bound principle is the following. Consider a minimization problem under certain constraints. Assume that an upper bound on the optimal value of the objective function is available. (This usually is the value of the objective function for the least feasible solution identified thus far). The first step is to partition the set of all feasible solutions into several subsets and, for each one, a lower bound is obtained for the value of the objective function of the solutions within that subset. Those subsets whose lower bounds exceed the current upper bound on the objective function are then excluded from further consideration. Now, of the remaining subsets one with the smallest lower bound is partitioned further into several subsets. Their lower bounds are obtained in turn and used as before to exclude some of these subsets from further consideration. From all of the remaining subsets, another one is selected for further partitioning etc. This process is repeated until a feasible solution is found such that the corresponding value of the objective function is no greater than the lower bound for any subset.

While solving an integer Programming problem the approach of Land and Doig is that for partitioning a .

feasible solution it adjoins the equality constraints such as $x_k = [t]$ and $x_k = [t] + 1$ etc. where $[t]$ is the largest integer less than or equal to t. This has a drawback that a large number of subsets may be required to be created and this number can not be predicted in advance. An improved partition which create, exactly two subsets from each eligible set is that if at a certain stage the optimal continuous solution has $x_k = t$ where t is not integral, then the first subset is created by introducing the inequality $x_k \leq [t]$ and the second is created by introducing $x_k \geq [t] + 1$. This variation in partitioning was given by Dakin, P (1965).

An integer solution to the convex programming problem stated in the chapter II could be detained by applying the above method. However, for the allocation problem of chapter III, which has been formulated as a programming problem of chapter II, a rule of thumb has been used for obtaining an integer solution.

1.4. Bayesian Estimation

The revelance of Bayes theorem to problems, arising in scientific investigations, in which inferences are made about parameter values, about which little is known apriori, is very great. Though, since the publication of Bayes work in 1763, there had been controversies about the acceptance and rejection of his work, recently the Bayes mode of reasoning has come up with great vigor.

Bayes Theorem

(i) Discrete Case . In an indeterminate practical situation a set of events $A_1, A_2, \ldots, A_k$ may occur. These events are mutually exclusive and exhaustive. Some other event A, is of particular interest. The probabilities, $P(A_i)$, $(i = 1, \ldots, k)$ of each of the A_i are known, as are the conditional probabilities, $P(A|A_i)$, $i = 1,2,\ldots, k$ of A given that A_i has occurred. Them the conditional (inverse) probabilities of any A_i, given that A has occurred, is given by

$$P(A_i|A) = \frac{P(A|A_i).P(A_i)}{\sum_{i=1}^{k} P(A|A_i)\,P(A_i)} . \qquad \ldots(1)$$

Rather than considering events A_i, we consider hypotheses $H_1, H_2, \ldots, H_k$ which constitute appropriate model for the practical situation. One and only one of these must be true, the event A becomes re-interpreted as an observed out come from the practical situation; it is the sample data. Prior to the observation, the probability, $P(H_i)$, that H_i is the appropriate model specification, is known for all i. These probabilities are the <u>prior probabilities</u> of the different hypotheses and constitute a secondary source of information. The probabilities, $P(A|H_i)$ $(i = 1,2,\ldots, k)$ of observing A, when H is the correct specification, are known also, these are simply the <u>likelihood</u> of the sample data.

We can reinterpret Bayes theorem as providing a means of updating, through use of sample data, our earlier state of knowledge expressed in terms of the prior probabilities, $P(H_i)$ $(i = 1,2,\ldots,k)$. The updated assessment is given by the <u>posterior probabilities</u>, $P(H_i|A)$ of the different hypotheses being true after we have utilized the further information provided by observing A to occur. This is the essence of Bayesian inference ; that the posterior probability of H_i given A is proportional to the product of the prior probability of H_i and the likelihood of A when H_i

is true. Prior information about the practical situation is in this way augmented by the sample data to yield a current probabilistic description of that situation, defined by the set of posterior probabilities.

(ii) Continuous Case

Suppose the $\underline{Y}' = (y_1, y_2, \ldots, y_n)$ is a vector of n observations whose probability distribution $p(\underline{y}|\underline{\theta})$ depends on the values of k parameters $\underline{\theta}' = (\theta_1, \ldots, \theta_k)$. Suppose also that $\underline{\theta}$ itself has a probability distribution. Then,

$$p(\underline{y}|\underline{\theta})p(\underline{\theta}) = p(\underline{y},\theta) = p(\theta|\underline{y})\ p(\underline{y})$$

Given the observed data $\underline{y}$, the conditional distribution of $\underline{\theta}$ is

$$p(\theta)|\underline{y}) = \frac{p(\underline{y}|\underline{\theta})\ p(\underline{\theta})}{p(\underline{y})} \qquad \ldots(2)$$

where
$$p(\underline{y}) = E\ p(\underline{y}|\underline{\theta}) = c^{-1} \begin{cases} \int p(\underline{y}|\underline{\theta})p(\underline{\theta})d\underline{\theta}, & \theta \text{ continuous} \\ \Sigma\ p(\underline{y}|\underline{\theta})p(\underline{\theta})d\underline{\theta}, & \theta \text{ discrete} \end{cases}$$

In this expression, $p(\theta)$, which tells in what is known about $\underline{\theta}$ without knowledge of data, is called the prior distribution of $\underline{\theta}$. Correspondingly, $p(\underline{\theta}|\underline{y})$, which tells us what is known about θ given knowledge of data, is called the posterior distribution of $\underline{\theta}$ given $\underline{y}$. Now given the data $\underline{y}$, $p(\underline{y}|\underline{\theta})$ may be regarded as a function not of $\underline{y}$ but of $\underline{\theta}$. It is called the likelihood function of θ for given $\underline{y}$ and can be written $\ell(\underline{\theta}|\underline{y})$. The Bayes formula thus becomes

$$p(\underline{\theta}|\underline{y}) = \ell(\underline{\theta}|\underline{y})\, p(\underline{\theta}). \qquad \ldots(3)$$

That is,

posterior distribution α likelihood x prior distribution.

The likelihood function $\ell(\underline{\theta}|\underline{y})$ plays a very important role in Bayes formula. It is the function through which data $\underline{y}$ modifies prior knowledge of θ, it can therefore be regarded as representing the information about θ coming from the data.

Sequential nature of Bayes Theorem

Bayes Theorem (3) provides a mathematical formulation of how previous knowledge may be combined with new knowledge. Indeed, the theorem allows us to continually update information about a set of parameters θ as more observations are taken.

Let $\underline{Y}_1$ be the initial sample of observations then

$$p(\underline{\theta}\,|\underline{y}_1) \;\alpha\; p(\theta)\; \ell(\underline{\theta}\,|\underline{y}_1) \qquad \ldots(4)$$

Suppose we have a second sample of observations $\underline{y}_2$, taken independently of the first sample, then

$$p(\underline{\theta}\,|\underline{y}_1, \underline{y}_2) \;\alpha\; p(\underline{\theta})\; \ell(\underline{\theta}\,|\underline{y}_1)\,.\,\ell(\underline{\theta}\,|\underline{y}_2)$$

$$\alpha\; p(\underline{\theta}\,|\underline{y}_1)\; \ell(\underline{\theta}\,|\underline{y}_2) \qquad \ldots(5)$$

The expression (5) is precisely of the same form as (4) except that $p(\underline{\theta}\,|\underline{y}_1)$, the posterior distribution for $\underline{\theta}$ given y_1, plays the role of the prior distribution for the second sample.

The point estimate of $\underline{\theta}$ is taken the value which minimizes the expected loss w.r.t. posterior distribution. Now if the loss is squared error, then the posterior mean is the optimum value of $\underline{\theta}$, [Rao, C.R. 1973], p.335. The expected loss for this estimate will be the variance of the posterior distribution.

1.5. Bayesian Posterior and Bayesian Preposterior Analysis

In Chapter IV and V optimum allocation is obtained using two different approaches, viz., posterior and preposterior analysis. In the Bayesian posterior analysis the decision is based on the combination of the prior information and the data information which are combined in the form of posterior distribution.

In preposterior analysis the decision is taken before actually performing the experiment. In this analysis we consider the prior expected value of all possible posterior expected utilities. According to Winkler, R.L. (1972) the preposterior analysis involves the potential posterior distribution following the proposed sample, which has not been observed yet ; it is just being contemplated.

In Chapter IV and V, in Bayesian posterior analysis, we make use of the first phase information both to allocate the second phase observation and also to estimate the value of over all mean for the j^{th} character, μ_j. Since all observations are incorporated into the estimator of μ_j i.e. $\sum_i \pi_i \bar{Y}_{ij}$, we choose the N_i to make the posterior distribution of μ_j as tight as possible, when π_i are known. This leads to the minimization of expected variance.

In preposterior analysis, the information from the first phase is used only to determine the best second phase allocation, but is not used for the estimation of μ_j. This estimator is chosen as $\Sigma\ \pi_i\bar{y}_{ij}$, and we make the distribution of this quantity as tight as possible, when π_i are known.

1.6. Non-Informative Prior and Conjugate Prior Distributions:

Non-Informatife Prior :

The solution to a inference problem is supplied by the posterior distribution $p(\underline{\theta}|\underline{y})$. The posterior distribution indicates what can be inferred about the parameter θ from the data $\underline{y}$ given a relevant prior state of knowledge represented by $p(\underline{\theta})$. The scientific investigation is performed only when the information supplied by the experiment is more precise than the information already available. This suggests that the likelihood must dominate the prior.

In general, prior which is dominated by the likelihood is one which does not change much over the region in which the likelihood is appreciable and does not assume large values out side the range. The prior distributions which have these properties are called locally uniform prior.

Suppose it is possible to express the unknown parameter θ in terms of a metric $\phi(\theta)$, so that the corresponding likelihood is data translated. This means that the likelihood curve for $\phi(\theta)$ is completely determined apriori except for its location which depends on the data yet to be observed. Then to say that we know little apriori relative to what the data is going to tell us, may be expressed by saying that we are almost equally willing to accept one value of $\phi(\theta)$ as another. This state of indifference may be expressed by taking $\phi(\theta)$ to be locally uniform, and the resulting prior distribution is called non-informative for $\phi(\theta)$ with respect to the data.

Jaffery rule . The prior distribution for a single parameter θ is approximately non-informative if it is taken proportional to the square root of Fishers Information measure. i.e. $p(\theta) \alpha \mathcal{J}^{1/2}(\theta)$.

In case of multi parameter $p'(\underset{\sim}{\theta}) \alpha |\mathcal{J}_n(\underset{\sim}{\theta})|$, where $\mathcal{J}_n(\theta)$ is the information matrix.

The firm of the non informative prior distribution depends upon the probability model of the observation.

Conjugate prior Distributions

Suppose the model for the practical situation under study is the family of distributions $\mathcal{P} = \{p_\theta(x), \theta \in \Omega\}$ indexed by the parameter θ, and the inference about θ is to be made. Suppose the prior distribution of θ is a member of some parametric family of distribution Q, with the property, in relation to $\mathcal{P}$, that the posterior distribution θ is also a member of Q. If this is so we say that Q is closed with respect to sampling from $\mathcal{P}$, or that Q is a family of conjugate prior distributions relative to $\mathcal{P}$.

Chapter II

A procedure For Convex Programming With Multiple Objective Functions

2.1. Introduction

There are many situations in which a system operates under the assumption of limited wealth and requires to accomplish multiple goals. The various ~~golas~~ goals depend upon the components parts of the system differently. A basic problem is to allocate the limited wealth among the components in such a way that all the goals are achieved to a maximum satisfaction.

If a precise weight can be assigned to each goal then a unique objective function can be defined by taking their weighted linear combination, Roy, B (1971). In certain cases the precise weight may not be known but a hierarchy of the various objective functions may be given. A method for dealing with such situations has been discussed by Benayoun, R. and J Tergny (1969).

In the absence of the apriori knowledge of relative weights or even an hierarchy, a STEP method has been developed for linear programming with multiple objective functions in

Benayoun R., J. Menteolfier and J. Tergny (1971). Here we generalize the above method for a typical situation in which the various objective functions are some convex functions while the constraints are linear.

2.2. Statement of the Problem

Let us denote I = 1,2,...,n . It is required to find a vector $\underline{x} = (x_1,\ldots,x_n)$ lying in a feasible region D defined by a linear constraint

$$\sum_{i=1}^{n} c_i x_i \leq c \quad , \qquad \ldots(2.1)$$

with $c_i \geq o$, $i \in I$ and $c \geq o$,

and the bounds

$$o \leq d_i \leq x_i \leq b_i \qquad i \in I \qquad \ldots(2.2)$$

which minimize in some sense all the following functions :

$$F_j = \sum_{i=1}^{n} a_{ij}/x_i \;, \; j \in J, \qquad \ldots (2.3)$$

where $J = 1,2,\ldots p$, $a_{ij} \geq o$ for $i \in I$ and $j \in J$

Since for $j \in J$ we have $\dfrac{\partial^2 F_j}{\partial x_i^2} = 2\, a_{ij}\, x_i^{-3} \geq o$ for $x_i \geq o$, it follows that the functions in (3) are convex in the feasible region D.

2.3. Obtaining the Solution For Individual Objective Functions

Consider the problem from (2.1) to (2.3) for a single objective function say k^{th}. First we minimize

$$F_k = \sum_{i=1}^{n} \frac{a_{ik}}{x_i} \qquad \ldots (2.4)$$

subject to (2.1) only. Due to the condition $c_i \geq o$, $i=1,2,\ldots,n$ it is easy to see that the solution to this problem will be on the boundary of the feasible region, i.e. where

$$\sum_{i=1}^{n} c_i x_i = c \qquad \ldots (2.5)$$

For, if the solution was such that $\Sigma\, c_i x_i < c$, then one could increase an x_i without violating the constraint (2.1) and thereby decreasing the value of F_k.

Using Lagranges Multipliers the solution to the problem of minimizing F_k subject to (2.5) has been derived by Khan, S and A. Bari (1977) as

$$x_i^k = (c_i a_{ik})^{\frac{1}{2}}\, c/c_i \sum_{i=1}^{n} (c_i\, a_{ik})^{\frac{1}{2}}, \; i \in I \qquad \ldots (2.6)$$

Note that $x_i^k \geq o$, $i \in I$ as all the constants in the R.H.S. of (2.6) are non-negative.

It is, of course, possible to solve the problem of minimizing F_x subject to the constraints (2.1) and (2.2) by any convex programming procedure but having to include the bounds in (2.2) into the constraints set greatly increases the computational effort. In many situations, particularly those to be considered in this thesis, it may seem unlikely, or even essentially impossible, that these bounds could turn out to be binding (i.e. hold as equality) on the optimal solution. If these bounds represent resource availabilities, past experience may indicate that the amount available should be more than adequate where as the constraints (2.1) imposes the primary restrictions on the solution.

An alternative procedure that yields satisfactory solution is to first finding the optimal solution to (2.4) to (2.5) as given in (2.6) where the bounds (2.2) have been ignored. Then check whether or not this solution violates any of the bounds. If not, the solution is, of course, optimal for the problem (2.1) to (2.3). On the other hand, if the solution does voilate one or more of the bounds, then the voilating variables are set equal to the corresponding bounds and the procedure is repeated for the problem in the remaining variables.

Suppose that the solution in(2.6) is such that $x_u^k < d_u$ for $u \in U$ and $x_v^k > b_v$ for $v \in V$. We fix $x_u^k = d_u$, $u \in U$ and $x_v^k = b_v$, $v \in V$.

After fixing x_i^k for $i \in U + V$, the optimal solution to the problem (2.4) - (2.5) is obtained as

$$x_i^k = (c_i a_{ik})^{\frac{1}{2}} \left(c - \sum_{u \in U} c_u d_u - \sum_{v \in V} c_v b_v\right) / \ c_i \Sigma (c_i a_{ik})^{\frac{1}{2}}$$

$$\text{for} \quad i \in I-U-V \qquad \ldots(2.7)$$

With this solution we again return to test whether the remaining bounds are now satisfied. If some of the bounds are again voilated by the solution in (2.7) then we must repeat as earlier by fixing the voilating variables equal to the corresponding bounds. The process is repeated until x_i^k , $i \in I$ satisfy all the bounds in (2.2).

2.4. Finding a Compromise Solution

The solution x_i^k, $i \in I$, found in (2.7) is optimal for the k^{th} objective function of the set F_j, $j \in J$. We determine the p different solutions x_i^1, $i \in I, x_i^p, i \in I$. Let the corresponding minimum values of the objective functions be $m_1, \ldots, m_p$. An ideal solution of the problem (2.1) to (2.3) would have been the one at which $F_j = m_j$ for all $j \in J$. But such a solution is, most likely, not feasible. We define a compromise solution as a point which is at a minimax distance to the ideal solution. This is equivalent to solving the following convex programming problem, Zukhuvitsky, S.I., and L. I. Avdeyeva, (1966).

Minimise w

subject to $F_j(\underline{x}) - m_j \leq w, \quad j \in J,$

$$\sum_{i \in I} c_i x_i \leq c$$

and $$d_i \leq x_i \leq b_i \qquad i \in I.$$

By making the transformation $x_i = \dfrac{1}{X_i}$, $i \in I$ and $w = X_{n+1}$, the above problem reduces to

$$\text{Minimize } X_{n+1} \qquad \ldots(2.8)$$

$$\text{subject to } \Sigma\, a_{ij} X_j - m_j \leq X_{n+1} \quad j \in I \qquad \ldots(2.9)$$

$$\sum_{i \in I} \frac{c_i}{X_i} \leq c \qquad \ldots(2.10)$$

and $$B_i \leq X_i \leq D_i, \quad i \in I \qquad \ldots(2.11)$$

where $$B_i = \frac{1}{b_i} \quad \text{and} \quad D_i = \frac{1}{d_i} \quad .$$

For a particular value of X_{n+1} the constraints (2.9) - (2.11) may be viewed as drawn in Fig. 1

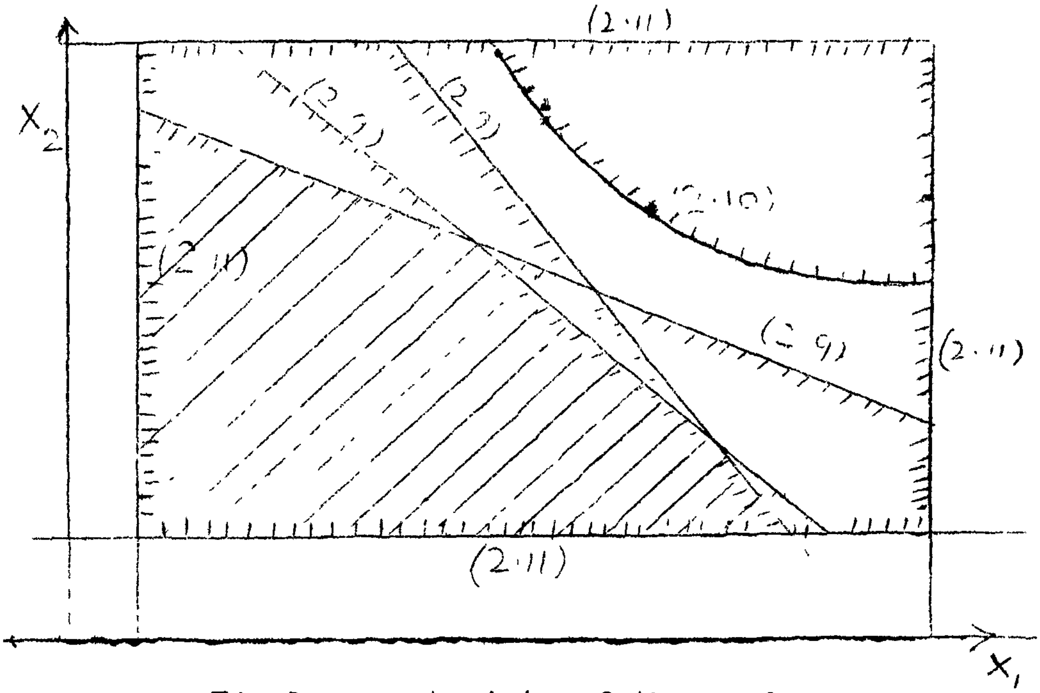


Fig.1 constraints of the problem (2.8) - (2.11)

The minimum of X_{n+1} will, of course, be non-negative For X_{n+1} = o, the points in the region defined by the constraints in (2.9) (which are clearly consistent, in the sense that there is at least one point, viz the origin, satisfying (2.9) since all the a_{ij}'s and m_j's are non-negative), will not satisfy the non-linear constraint(2.10). Any increase in X_{n+1} will amount to a displacement upwards of the linear constraints (2.9). A solution to the problem (2.8) to (2.11) may thus be obtain by moving the region defined by the constraints

(2.9), through changes in X_{n+1} to the extent that the region just touches the feasible region defined by the non-linear constraint (2.10).

To this end we proceed as follows

Fix the value of X_{n+1} as a small positive number $X_{n+1}^{(t)}$, where t stands for the t^{th} iteration in the following. Then solve the following convex programming problem

$$\text{Minimize} \quad G = \sum_{i \in I} c_i / x_i \qquad \ldots(2.12)$$

$$\text{subject to} \quad \sum a_{ij} X_i \leq m_j + X_{n+1}^{(t)} \;, \; j \in J \qquad \ldots(2.13)$$

$$\text{and} \quad B_i \leq X_i \leq D_i \qquad i \in I \qquad \ldots(2.14)$$

Let the minimum value of the objective function for the problem (2.12) - (2.14) be G_t.

If $G_t - c \neq o$, it is clear that a feasible solution for the problem in (2.8) - (2.11) has not been attained by the value of X_{n+1} as $X_{n+1}^{(t)}$. We want to determine a value of X_{n+1} for which the optimum value of G differs from c by less than some pre-assigned number say ϵ. A usual one dimensional search method is employed to determine the required value of X_{n+1} as follows.

We put $X_{n+1}^{(t+1)} = X_n^{(t)} + \delta^{(t)}$ where $\delta^{(t)}$ is an arbitrary number greater than zero or less than zero according as $G_t - c > o$ or $< o$. For starting we may take $\delta^{(1)} = (G_1 - c)/_c$. Then solve the problem (2.12) to (2.14) with $X_{n+1}^{(t)}$ replaced by $X_{n+1}^{(t+1)}$. This process is continued with $\delta^{(t)} = 2\delta^{(t-1)}$ and $X_{n+1}^{(t+1)} = X_{n+1}^{(t)} + \delta^{(t)}$ for t = 2,3,... .until for t = r, say, the sign of $G_r - c$ changes for the first time. Then for t = r+1 we take $\delta^{(r+1)} = -\frac{\delta^{(r)}}{2}$ and $X_{n+1}^{(r+2)} = X_{n+1}^{(r+1)} + \delta^{(r+1)}$. At further steps, say ℓ^{th}, $\delta^{(\ell)} = -\frac{\delta^{(\ell-1)}}{2}$ if G-c has changed the sign at ℓ-1 step and $\delta^{(\ell)} = \frac{\delta^{(\ell-1)}}{2}$ otherwise. The process terminates when $|G_t - c| < \epsilon$ for some t.

2.5. Further Improvement In the Solution

It may happen that after observing the minimax solution by solving the problem (2.12) - (2.14) one feels that the solution is not quite pleasing for some of the objective functions F_j in (2.3). In this case one may improve his solution for some objective functions on the cost of tolerating a certain amount of increase in the values of the others, which were satisfactory at the compromise solution obtained earlier.

If some of the $F_j(\underline{x})$ are satisfactory and others are not then for allowing an improvement in the unsatisfactory ones a certain amount of increase must be accepted from m_j corresponding to the satisfactory $F_j(\underline{x})$. Let the index of the objective to be relaxed be j^*. An information on the selection of the index j^* can be obtained by performing a sensitivity analysis for the problem in (2.8) to (2.11).

Let Δm_{j*} be the amount of increase accepted in m_{j*}. Incorporating this change in the problem (2.12) to (2.14) an improved compromize solution is obtained by solving the problem (by the method described earlier) :

$$\text{Minimize} \quad \sum_{i \in I} c_i/X_i \qquad \ldots(2.15)$$

$$\text{subject to} \quad \sum_{i \in I} a_{ij} X_i \leq m_j + X_{n+1}^{(t)}, \; j \neq j^* \in J \qquad \ldots(2.16)$$

$$\sum_{i \in I} a_{ij} X_i \leq (m_{j*} + \Delta m_{j*}) + X_{(n+1)}^{(t)} \qquad \ldots(2.17)$$

and

$$B_i \leq X_i \leq D_i \qquad i \in I \qquad \ldots(2.18)$$

If further improvement are needed then the above process may be repeated by solving the problem (2.15) to (2.18) after making the fresh adjustment in the satisfactory $F_j(\underline{x})$.

2.6. Obtaining An Integer Solution

An exact compromise integer solution could be obtained by applying the branch and bound procedure using Dakin's approach given in Dekin, R (1965) to the problem (2.12) to (2.14). It may be noted that our stretegy in the procedure used in Section 4 to solve the problem (2.12) to (2.14) was to keep $|G-c|$ at the minimum level. We give in the following an easily available method in which the budgetory constraint (2.10) is not allowed to be voilated by the final integer solution.

We arrange c_i, $i \in I$ such that $c_{(1)} \geq c_{(2)} \geq \dots c_{(n)}$ Denote the corresponding x_i by $x_{(i)}$, $i \in I$. The following step is repeated for $j = 1,2,\dots, n-1$. At j^{th} step compute

$$S_1 = \left| \sum_{i=1}^{j-1} c_{(i)} x^*_{(i)} + c_{(j)} [x_{(j)}] + \sum_{i=j+1}^{n} c_{(i)} x_{(i)} - c \right|$$

and

$$S_2 = \left| \sum_{i=1}^{j-1} c_{(i)} x^*_{(i)} + c_{(j)} ([x_{(j)}]+1) + \sum_{i=j+1}^{n} c_{(i)} x_{(i)} - c \right|$$

where $|x|$ represents the integral part of x. Fix $[x_{(j)}] = x^*_{(j)}$ if $S_1 \leq S_2$. Otherwise fix $[x_{(j)}]+1 = x^*_{(j)}$.

For $j = n$ if $S_2 \neq o$, then we should fix $|x_n| = x_n^*$ even if $S_1 > S_2$. This is done for maintaining $\Sigma\, c_{(i)}\, x_i^* \leq c$. x_i^*, $i \in I$ constitutes an approximate compromize integer solution to the problem (2.1) to (2. 3).

Chapter III

Optimum Allocation in Multiple Character Stratified Random Sampling In the Presence of Prior Information

3.1. Introduction. Optimum allocation in stratified random sampling for a single estimation character has been discussed by Ericson (1965). He also discussed the case when 'p' population characteristics are to be estimated, but only under the assumption that the various strata are sufficiently similar with respect to (p-1) characteristics.

Ahsan, J and S.U.Khan (1977) have proposed a solution procedure for the case involving p characters in which no assumption about the similarity of strata is made with respect to the various characters. The situation has been formulated as to minimize the cost of the survey subject to desired precisions assigned to the posterior variances of the means of the different characters. Here we consider the situation in which the budget is fixed and the precisions for the various characters is to be maximized.

For a fixed budget an allocation advantageous to one character in the sense of maximizing the precision may produce unhappy results for others. A unique objective function can be defined when precise weight is known for each character in the survey, Roy (1971). In the absence of such apriori knowledge of relative weights the problem can not be exactly transformed to give a unique objective function and hence a compromising solution is obtained for the multiple objective functions by using the procedure described in Chapter II.

3.2. Formulation of the Allocation Problem

We assume that the strata boundaries are fixed in advance and the sample units are chosen independently in different strata. Let there be p characters under study and k different strata. Let Y_{ij} be the variable corresponding the j^{th} character in the i^{th} stratum where $i \in I = 1, \ldots, k$ and $j \in J = 1, \ldots, p$. We assume that Y_{ij}'s are independently normally distributed with means θ_{ij}'s and known variance σ^2_{ij}'s, $i \in I$, $j \in J$. A simple random sample of n_i units is drawn from the i^{th} stratum. Let the observations be $y_{ij1}, y_{ij2}, \ldots, y_{ijn_i}$. Then the likelihood function of θ_{ij}, given the observations on Y , $L(\theta_{ij}|Y)$, is proportional to

$$\prod_{i=1}^{k} \left(\frac{1}{2\pi\,\sigma_{ij}}\right)^{n_i} \exp\left[-\frac{1}{2\,\sigma_{ij}^2} \sum_{=1}^{n_i} (y_{ij} - \theta_{ij})^2\right]. \quad \ldots(3.1)$$

Since

$$\sum_{=1}^{n_i} (y_{ij} - \theta_{ij})^2 = \sum_{=1}^{n_i} (y_{ij} - \bar{y}_{ij})^2 + n_i(\theta_{ij} - \bar{y}_{ij})^2, \quad \ldots(3.2)$$

we get

$$L(\theta_{ij} | \underline{Y}) \;\alpha\; \prod_{i=1}^{k} \exp \left| -\frac{1}{2} \left(\frac{\theta_{ij} - \bar{y}_{ij}}{\sigma_{ij}/\sqrt{n_i}}\right)^2 \right|. \quad \ldots(3.3)$$

We assume that prior information on θ_{ij}, available before drawing the sample is in the form of apriori distribution of θ_{ij} as normal with mean μ_{oij} and variance $^2_{oij}$, called conjugate prior, Raiffa and Shlaiffer (1961). Then the posterior distribution of θ_{ij} obtained by combining the likelihood function (3.3) with a Normal prior $N(\mu_{oij}, \sigma^2_{oij})$ is the Normal distribution $N(\mu_{1ij}, \sigma^2_{1ij})$ Box and Tio (1973),

where

$$\mu_{1ij} = \frac{1}{w_{oij} + w_{ij}} \left[w_{oij}\, \mu_{oij} + w_{ij}\, \bar{y}_{ij}\right] \quad \ldots(3.4)$$

and

$$\frac{1}{\sigma^2_{1ij}} = w_{oij} + w_{ij} \,. \quad \ldots(3.5)$$

w_{oij}, being the reciprocal of the variance of the apriori distribution of θ_{ij}, is a measure of apriori information on θ_{ij}, and $w_{ij} = \frac{n_i}{\sigma^2_{ij}}$ is a measure of information in the sample of n_i from the i^{th} stratum. It could be mentioned that the combined information in the i^{th} stratum is the sum of the apriori and the sample information.

Our object is to estimate the population mean $\theta_j = \sum_{i=1}^{k} \pi_i \, \theta_{ij}$ for the j^{th} character, where π_i is the size of the i^{th} stratum. Bayes estimate of θ_{ij} is μ_{1ij} given in (3.4).

As the sampling in each stratum is independent, we have the posterior variance of θ_j, i.e. $V(\theta_j/\underline{Y})$ as

$$V(\theta_j | \underline{Y}) = \sum_{i=1}^{k} \pi_i^2 \ \sigma^2_{1ij} \qquad j \in J \qquad \ldots(3.6)$$

Let c be total available budget for the survey and c_i be the cost of completely measuring a unit in the i^{th} stratum. Then we must have $\sum_{i=1}^{k} c_i n_i \leq c$.

The problem of optimum allocation is to find n_i's such that the posterior variance of θ_j for the j^{th} character given the observations, is minimized for all while the budgetary restriction is not voilated.

Posterior variance of θ_j, given the observations, is

$$V(\theta_j \mid \underline{Y}) = \sum_{i \in I} \pi_i^2 \, \sigma^2_{1ij} = \sum_{i=1}^{k} \frac{\Sigma\, \pi_i^2}{w_{oij} + w_{ij}} = \Sigma \frac{\pi_i^2}{\frac{1}{\sigma^2_{oij}} + \frac{n_i}{\sigma^2_{ij}}}$$

$$= \sum_{i=1}^{k} \pi_i^2 \quad {}^{2}_{ij} \left(n_i + \frac{\sigma^2_{ij}}{\sigma^2_{oij}}\right)^{-1}$$

$$= \sum_{i=1}^{k} \pi_i^2 \, \frac{\sigma^2_{ij}}{n_i} \left(1 + \frac{\sigma^2_{ij}}{n_i \, \sigma^2_{oij}}\right)^{-1}.$$

In scientific experiments the likelihood dominates the prior. Firstly because a scientific investigation is not usually undertaken unless information supplied by the experiment is likely to be more precise than information already available. Secondly, if an experimenter holds prior strong belief about the value of the parameter, then in reporting his result it would be appropriate and most convincing to analyse the data against a reference prior which is dominated by the likelihood. In the light of the above reasons w_{oij} will be negligible compared to w_{ij}. Thus the higher order terms in the expansion of $\left(1 + \frac{\sigma^2_{ij}}{n_i \; {}_{oij}}\right)^{-1}$ may be neglected and we obtain approximately

$$\Sigma\, \pi_i^2 \, \frac{\sigma^2_{ij}}{n_i} \left(1 + \frac{\sigma^2_{ij}}{n_i \; {}^{2}_{oij}}\right)^{-1} \simeq \Sigma\, \pi_i^2 \, \frac{\sigma^2_{ij}}{n_i} = V_j, \text{ say}.$$

Consequently, our problem of choosing the sample numbers n_i, $i \in I$, to the various strata, reduces to

$$\text{Minimise} \quad V_j = \sum_{i \in I} \frac{v_{ij}}{n_i}, \quad j \in J \qquad \ldots(3.7)$$

$$\text{under the constraint} \quad \sum_{i=1}^{k} c_i n_i \leq c \qquad \ldots(3.8)$$

where $v_{ij} = \pi_i^2 \quad {}^{2}_{ij}$.

Since one should take atleast one observation from each stratum we must have a further restriction.

$$n_i \geq 1, \quad i \in I \qquad \ldots(3.9)$$

3.3. Solution Procedure

The problem in (3.7) to (3.9) has the form of the convex programming problem with multiple objective functions (2.1) to (2.3) treated in Chapter 2. The procedure described there in sections 4 and 5 may be used for finding a compromise solution. Further, it is known, Cochran, W.G. 1977 that the cufves of precision are flat in the vicinity of optimum allocation. So the integer solution procedure of section 6 of Chapter 2 may be applied which does not let the budgetary constraints to be voilated while the objective functions remain close to the optimal point.

3.4. Numerical Example

A sample survey is to be conducted for studying five different characters with a total available budget of 160 units. The population is infinite which has been partitioned into four large sized strata. The known constants v_{ij} and the costs of completely enumerating a unit in the various strata are given in the following table

Characters

	v_{ij}	1	2	3	4	5	c_i
	1	3.4	5.8	2.4	1.8	2.9	
	2	3.9	1.6	4.8	2.8	5.9	
Strata	3	2.2	4.4	1.0	5.7	3.6	
	4	5.0	2.2	3.9	1.3	4.8	

The allocation problem as formulated in (3.7) to (3.9) may be stated as follows :

$$\text{Minimize} \quad \sum_{i=1}^{4} \frac{v_{ij}}{n_i}, \quad j = 1,2,\ldots, 5 \qquad \ldots(3.10)$$

$$\text{subject to} \quad 2n_1 + 3n_2 + n_3 + 2n_4 \leq 160 \qquad \ldots(3.11)$$

$$\text{and} \qquad n_i \geq 1 \quad i = 1,2,\ldots, 4 \qquad \ldots(3.12)$$

In the following table the j^{th} column gives the allocation which is optimal for the j^{th} character (obtained by using 2.6) and the total cost incurred by the allocation.

The last row represents the corresponding minimum variances.

n_i^j	1	2	3	4	5
1	19.4	27.7	17.9	17.3	16.5
2	16.9	11.9	20.7	17.6	19.3
3	22.4	34.3	16.3	43.6	26.2
4	23.8	17.1	22.0	14.5	21.4
$\Sigma c_i n_i^j$	159.5	159.6	158.3	160.0	159.9
m_j	0.71	0.59	0.59	0.49	0.84

For finding the compromise solution the following problem is to be solved (see the problem (2.12) to (2.14) :

$$\text{Minimise} \quad \frac{2}{x_1} + \frac{3}{x_2} + \frac{1}{x_3} + \frac{2}{x_4} \qquad \ldots(3.13)$$

$$\text{subject to} \quad \Sigma\, v_{ij}\, x_i \leq m_j + x_5^{(1)}, j = 1,2,\ldots,5 \quad ..(3.14)$$

$$\text{and} \quad o \leq x_i \leq 1,\; i = 1,2,\ldots, 4 \qquad \ldots(3.15)$$

Let us fix initially $x_5^{(1)} = 0.5$, the solution to the problem (3.13) to (3.15) obtained by using (2.7) is

$$x_1 = 0.090,\ x_2 = 0.084,\ x_3 = 0.059,\ x_4 = 1.410.$$

The corresponding value of the objective function (3 13) is 101.4 which is much less than 160.

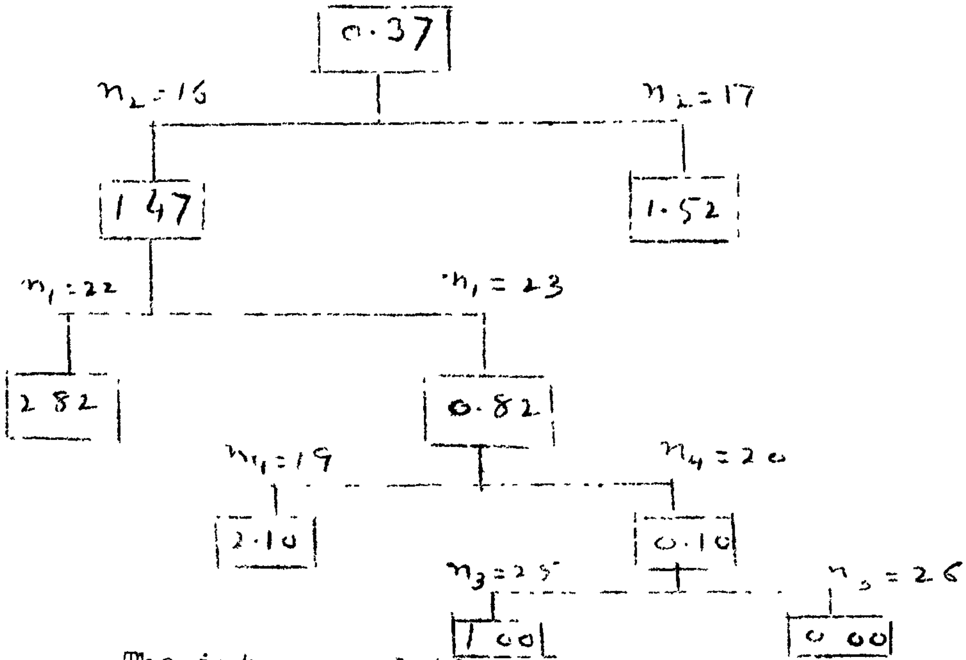
Next we solve the problem (3.13) to (3.15) with $x_5^{(1)}$ replaced by $x_5^{(2)} = x_5^{(1)} + \frac{101.4-160}{160}$. The new value of the objective function still falls short of 160. We continue the iterations. When $x_5 = 0.0469$ the solution of the new problem in (3.13) to (3.15) is

$$x_1 = 0.0441,\ x_2 = 0.0609,\ x_3 = 0.0386,\ x_4 = 0.0509,$$

and the corresponding value of the objective function is 159.63. This value is close to 160. The non-integer solution obtained from these values is

$$n_1 = 22.67,\ n_2 = 16.41,\ n_3 = 25.90,\ n_4 = 19.64.$$

An integer solution is obtained by applying the procedure explained in section 2.6 . The various steps are represented in the following figure. The numbers in the rectangles are the values of $| \Sigma\, c_i n_i - c |$



The integer solution is thus

$$n_1 = 23, \; n_2 = 16, \; n_3 = 26, \; n_4 = 20.$$

Chapter IV

ltiple Character Stratified two ~~space~~ phase Sampling Using Non informative Prior

4.1. Introduction.

Bayesian stratified two phase sampling was studied by Draper, N.R. and I Guttman (1968). They obtained the results for the optimum allocation of the sample sizes in different strata at the second phase using the non-informative prior before the first phase. A definite proportion of the total fixed budget was allocated for the sampling in the first phase and the rest for the second phase, and the posterior variance of the mean of the only character under study was minimized. Here we consider the case when more than one characters, say p, are studied. Let us first assume that these p characters are independently distributed. The case where the different characters are not independently distributed has been dealt with in sec 4.5.

The posterior variances of the overall mean for each of the characters are considered for minimization. The problem has been formulated as a convex programming problem with multiple objective functions which also can be solved by using the method developed in Chapter II.

4.2. Description of the Problem

The population is divided into k strata and the strata boundaries are fixed. The size of the i^{th} stratum is π_i $(i = 1,2,\ldots, k)$. Let μ_{ij} be the i^{th} stratum mean for the j^{th} character. The objective of the sampling is to estimate $\mu_j = \sum_i \pi_i \mu_{ij}$ the population mean for the j^{th} character.

For j^{th} character in the i^{th} stratum let X_{ij} and Y_{ij} $(i = 1,2,\ldots, k)$. be normally distributed variable with unknown mean μ_{ij} and unknown variance σ^2_{ij}, where the letters X and Y denotes the variables for the first phase and for the second phase respectively.

We assume that the prior information available before the first phase sampling can be represented by independent locally uniform prior distribution of μ_{ij} and σ^2_{ij}, that is,

$$p(\mu_{ij},\sigma^2_{ij}) = p(\mu_{ij})p(\sigma^2_{ij}) \qquad d\,\mu_{ij} \;\; \frac{d\sigma^2_{ij}}{\sigma^2_{ij}} \qquad \ldots(4.1)$$

Let c_i be the cost of taking one observation from the i^{th} stratum and c be the total cost of the survey. Let n_i and N_i be the sample sizes at the first phase and at the second phase respectively in the i^{th} stratum. A fraction βc of the total budget, where $o < \beta < 1$, is alloted for the first phase and $(1-\beta)c$ for the second phase. We have

$$\Sigma\ c_i n_i = \beta c,\ \Sigma\ c_i N_i = (1-\beta)c \qquad \ldots(4.2)$$

The problem is to choose N_i, while n_i is fixed, so that the posterior variances for the over all means μ_j, $j = 1,2,\ldots,p$, are minimized while the conditions in (4.2) hold.

4.3. Posterior Analysis

(i) First we consider the case where the strata sizes are known.

The sample size n_i in the first phase in the i^{th} stratum is already specified. The observations which are in the first phase of a two phase sampling procedure from the i^{th} stratum are $x_{ij1}, x_{ij2}, \ldots, x_{ijn_i}$. The likelihood is, Box and Tio (1973), p. 63.

$$\prod_{i=1}^{k} (\sigma_{ij}^{2})^{-\frac{n_i}{2}} (2\pi)^{-\frac{n_i}{2}} \exp -\frac{n_i(\bar{x}_{ij}-\mu_{ij})^2+(n_i-1)s_{ij}^2}{2\sigma_{ij}^2}, \quad ..(4.3)$$

where $n_i\ \bar{x}_{ij} = \sum_{\ell=1}^{n_i} x_{ij\ell}$, $(n_i-1)s_i^2 = \sum_{\ell=1}^{n_i}(x_{ij\ell} - \bar{x}_{ij})^2$.

The joint posterior distribution of μ_{ij} and $\sigma_{ij}^2 (i = 1,2,\ldots,k)$ after the first phase observations have been taken is proportional to the product of (4.1) and (4.3), i.e.,

$$\prod_{i=1}^{k} p(\mu_{ij},\sigma_{ij}^2 \mid x_{ij1},x_{ij2},\ldots,x_{ijn_i}) \propto$$

$$\prod_{i=1}^{k} (\sigma_{ij}^{2})^{-\frac{n_i}{2}-1} \exp -\frac{n_i(\bar{x}_{ij}-\mu_{ij})^2+(n_i-1)s_{ij}^2}{2\sigma_{ij}^2} \quad . \ldots(4.4)$$

The observations in the second phase are $y_{ij1}, y_{ij2},\ldots, y_{ijN_i}$. Then proceeding in the same way as we obtained the joint posterior after the first phase, we obtain the joint posterior after the second phase, using (4.4) as the prior before the second phase.

$$\prod_{i=1}^{k} p(\mu_{ij},\sigma_{ij}^2 \mid x_{ij1}, x_{ijn_i}, y_{ij1},y_{ij2}.. y_{ijN_i})$$

$$\propto (\sigma_{ij}^{2})^{-\frac{1}{2}(N_i+n_i+2)} \exp\left(-\frac{Q_{ij}}{2\sigma_{ij}^2}\right) \quad \ldots(4.5)$$

where $Q_{ij} = n_i(\bar{x}_{ij}-\mu_{ij})^2 + N_i(\bar{y}_{ij}-\mu_{ij})^2 + (n_i-1)s_{ij}^2 + (N_i-1)w_{ij}^2$

$$= (N_i+n_i)(\mu_{ij}-\bar{Y}_{ij})^2 + (SS)_{ij} \qquad \ldots(4.6)$$

where $(N_i-1)w_{ij}^2 = \sum_{\ell=1}^{N_i} (y_{ij\ell} - \bar{y}_{ij})^2$

$$(SS)_{ij} = \frac{n_i N_i}{N_i+n_i}(\bar{x}_{ij} - \bar{y}_{ij})^2 + (n_i-1)s_{ij}^2 + (N_i-1)w_{ij}^2$$

=Corrected total S.S. for the total sample in the i^{th} stratum.

$$\bar{Y}_{ij} = \frac{n_i \bar{x}_{ij} + N_i \bar{y}_{ij}}{n_i + N_i} \qquad \ldots(4.7)$$

The marginal distribution of μ_{ij}, obtained by integrating out σ_{ij}^2 comes after little manuplutation as

$$\prod_{i=1}^{k} (\mu_{ij} | \underline{x}_{ij}, \underline{y}_{ij}) \propto \prod_{i=1}^{k} \left\{ 1 + \frac{T_{ij}^2}{N_i + n_i} \right\}^{-\frac{N_i+n_i}{2}}, \qquad \ldots(4.8)$$

where $T_{ij}^2 = \dfrac{(N_i+n_i)(N_i+n_i-1)(\mu_{ij}-\bar{Y}_{ij})^2}{(SS)_{ij}}$.

As T_{ij} has a posterior t-distribution with (N_i+n_i-1) d.f. we have

$$E(T_{ij}) = 0 \quad \text{i.e.,}$$

$$E(\mu_{ij}) = \bar{\bar{Y}}_{ij} ,$$

and $$E(T_{ij}^2) = \frac{N_i + n_i - 1}{N_i + n_i - 3} .$$

It follows that posterior variance of μ_{ij} after the first and second phase observations have been taken is

$$V(\mu_{ij}) = E(\mu_{ij} - \bar{\bar{Y}}_{ij})^2 = \frac{(SS)_{ij}}{(N_i+n_i-3)(N_i+n_i)} \qquad \ldots(4.9)$$

Now the problem of optimum allocation is to choose N_i's such that the posterior variance of $\tilde{\mu}_j$, i.e., $V(\sum_{i=1}^{k} \pi_i \mu_{ij}) = \Sigma\, \pi_i^2\, V(\mu_{ij})$, is minimum subject to (4.2). But the expression for $V(\mu_{ij})$ in (4.9) involves $(SS)_{ij}$ which itself depends on the second phase observations. So replace $(SS)_{ij}$ by $E(SS)_{ij}$, where the expectation is taken over the future distribution of second phase observations, given the first phase observations. Using the result of section 4 for pre-posterior analysis we have

$$E(SS)_{ij} = \frac{(N_i+n_i-3)(n_i-1)}{(n_i-3)\ s_{ij}^2} \qquad \ldots(4.10)$$

Substituting for $E(SS)_{ij}$ in $\Sigma\ \pi_i^2\ V(\mu_{ij})$ for $(SS)_{ij}$ and simplifying we have

$$E_y\ V(\mu_j) = \Sigma\ \pi_i^2\ u_{ij}^2 / \ (N_i+n_i), \qquad \ldots(4.11)$$

where $u_{ij}^2 = \dfrac{(n_i-1)s_{ij}^2}{(n_i-3)}$.

Now put $(N_i+n_i) = x_i$.

Then $E_y\ V(\mu_j) = \Sigma\ \pi_i^2\ \dfrac{u_{ij}^2}{x_i}$.

Consequently the problem reduces to as

$$\text{Minimise} \quad \Sigma\ \frac{v_{ij}}{x_i}\ ,\ j \in J$$

subject to $\quad \Sigma\ c_i\ x_i \ \leq\ c$

and $\quad x_i \geq n_i,\quad i \in I$

where $v_{ij} = \pi_i^2\ u_{ij}^2$. The restriction $x_i \geq n_i$ has been imposed to avoid negative N_i, which is in the form of (2.12) to (2.14).

(ii) Now consider the case where the strata sizes π_i are unknown.

When the π_i are unknown, we assume the apriori distribution of π_i as the Drichlet distribution

$$p(\pi_1,\pi_2,\pi_k)=\frac{\Gamma(\nu_1+\nu_2+\ldots+\nu_k)}{\prod_{i=1}^{k}\Gamma(\nu_i)}\,\pi_1^{\nu_1-1}\,\pi_2^{\nu_2-1}\,\pi_k^{\nu_k-1},\ldots(4.12)$$

where $\sum_{i=1}^{k}\pi_i = 1$. The reason for taking Dirichlet as the apriori is that it includes as special cases, the locally uniform prior (all $\nu_i=1$), invariant prior (all $\nu_i=\frac{1}{2}$) and is the appropriate family of conjugate priors, Raiffa and Schlaiffer, (1961). For proportion the prior distribution of μ_{ij} and σ^2_{ij} are the same as before.

The sampling at the first phase can not be taken according to proportional allocation which is of course better than arbitrary allocation as the strata sizes are unknown. To avoid this difficulty a portion γc (the maximum budget meant for the first phase) where $\gamma < \beta$, of the total budget decides about the first phase total sample size as $\left[\frac{\gamma c}{(\max c_i)}\right]$. A random sample of this size is drawn from the population and say n_i of them fall in the i^{th} stratum. Let $\Sigma\, n_i\, c_i = \beta c$.

The likelihood function is proportional to the product of (4.3) and the multinomial likelihood of n_i's, which is

$$\prod_{i=1}^{k} \frac{1}{(\sigma_{ij})^{n_i}} \exp\left[-\frac{n_i(\bar{x}_{ij}-\mu_{ij})^2+(n_i-1)s_{ij}^2}{2\sigma_{ij}^2}\right] \frac{(n_1+..+n_k)!}{n_i!}\pi_i^{n_i} \qquad \ldots(4.13)$$

The posterior distribution of μ_{ij}, σ_{ij}^2 and π_i's after the first phase is the product of priors of μ_{ij} and σ_{ij}^2 i.e. (4.1) and (4.12) with (4.13). This is used as a prior in the second phase. Combining the likelihood at the second phase with this posterior distribution of μ_{ij}, σ_{ij}^2 and π_i's after the first phase, we obtain the posterior distribution of μ_{ij}, σ_{ij}^2 and π_i given the first phase, second phase observations and the n_i's. It can be shown that if $V(\mu_j)$ is evaluated and its expectation taken over the future distribution of $\underline{Y}$, it reduces to

$$\Sigma\, \frac{D_i^2\ u_{ij}^2}{N_i+n_i} + \text{terms not involving } N_i, \qquad \ldots(4.14)$$

where $D_i = \dfrac{n_i+\nu_i}{\sum\limits_{i=1}^{k}(n_i+\nu_i)} = E(\pi_i)$, and $u_{ij}^2 = \dfrac{(n_i-1)s_{ij}^2}{(n_i-3)}$.

Which is in the form of (4.11).

4.4 Preposterior Analysis

We consider the case where the strata sizes are known.

The joint posterior distribution after the first phase, i.e. (4.3), is used (by integrating out in the parameters) as the future distribution of sample means $\bar{y}_{ij}$ at the second phase. The joint distribution of the means $\bar{y}_{ij}$'s given μ_{ij} and σ_{ij}^2 is

$$p(\bar{y}_{ij},\bar{y}_{2j},\ldots,\bar{y}_{kj}\,|\mu_{ij},\sigma_{ij}^2)=\frac{1}{\sigma_{ij}}\sqrt{\left(\frac{N_i}{2\pi}\right)}\exp\left\{-\frac{N_i(\bar{y}_{ij}-\mu_{ij})^2}{2\sigma_{ij}^2}\right\} \qquad \ldots(4.15)$$

The joint distribution of $\bar{y}_{ij}$, μ_{ij}, σ_{ij}^2 after the first phase, with (4.15) which after simplification comes as

$$\prod_{i=1}^{k} p(\bar{y}_i,\ \mu_{ij},\sigma_{ij}^2) =$$

$$\prod_{i=1}^{k}\frac{N_i^{\frac{1}{2}}}{(\sigma_{ij}^2)^{\frac{(n_i+3)}{2}}}\exp\left[-\frac{1}{2\sigma_{ij}^2}\left\{(\bar{y}_{ij}-\mu_{ij})^2N_i+\right.\right.$$

$$\left.\left.+n_i(\bar{x}_{ij}-\mu_{ij})^2+(n_i-1)s_{ij}^2\right\}\right]. \qquad \ldots(4.16)$$

Integrating out the parameters of (4.16), the preposterior distribution of the means $\bar{y}_{ij}$ is proportional to the product of t distributions, i.e.

$$\prod_{i=1}^{k} p(\bar{y}_{ij}) \propto \prod_{i=1}^{k} \left(1+ \frac{t_{ij}^2}{\nu_i}\right)^{-\frac{1}{2}(\nu_i+1)}$$

where $t_{ij}^2 = n_i\ N_i(\bar{y}_{ij} - \bar{x}_{ij})^2 / s_{ij}^2\ (N_i+n_i)$

and $\nu_i = n_i - 1.$

From the conditional mean and variance of t-distribution we have

$$E(\bar{y}_{ij}) = \bar{x}_{ij} \qquad \ldots (4.17)$$

and $$V(\bar{y}_{ij}) = (n_i-1)\left(\frac{N_i+n_i}{N_i\ n_i}\right)\frac{s_{ij}^2}{n_i-3}\ . \qquad \ldots (4.18)$$

Since aposteriori expectation of $\Sigma\pi_i\mu_{ij}$ after the first phase and preposterior before the second phase expectation of $\sum_i \pi_i\ \bar{y}_{ij}$ are each $\Sigma\ \pi_i\bar{x}_{ij}$, this suggests that $\sum_i \pi_i\ \bar{y}_{ij}$ is a preposterior estimator of $\mu_j=\Sigma\pi_i\mu_{ij}$. The estimate $\sum_i \pi_i\ \bar{y}_{ij}$ has variance, from (4.18), as equal to

$$\Sigma A_{ij}(1+\frac{n_i}{N_i}) = \sum_i A_{ij}+ \sum_i A_{ij} \frac{n_i}{N_i}, \qquad \ldots(4.19)$$

where $A_{ij} = \pi_i^2 \ (1-\frac{1}{n_i}) \ (\frac{s_{ij}^2}{n_i-3})$.

The first term of (4.19) is known guantity. We minimize (4.19) w.r.t N_i subject to the budget restriction $\Sigma c_i N_i \leq (1-\beta)c$.

4.5. Posterior Analysis in the case where the characters are correlated.

Let the row vector variable $\underline{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})'$ be normally distributed with mean vector $\underline{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})$ and covariance matrix $\Sigma_i = (\sigma^{i\,2}_{jj'})$, where $\sigma^{i\,2}_{jj'}$ is the covariance of X_{ij} and $X_{ij'}$.

Samples of sizes n_i and N_i are drawn from $\underline{X}_i$ and $\underline{Y}_i \sim N_p(\underline{\mu}_i, \Sigma_i)$ where $\underline{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})'$, $\underline{X}_i = (X_{i1}, \ldots, X_{ip})'$ and Σ_i is the covariance matrix of the p characters in the i^{th} stratum. Prior information available on $\underline{\mu}_i$ and Σ_i are locally uniform and invariant prior distributions.

As Σ_i is +ve definite, there exists an orthogonal p x p matrix, say P, such that $P \Sigma_i P' = \Lambda$, where Λ is a diagonal matrix, Rao, C.R. (1973).

Make the transformation $\underline{Z}_i = P\underline{X}_i$. Then the elements of $\underline{Z}_i$ will be independently normally distributed. Starting with the elements of $\underline{Z}_i$ and proceeding in the same way as in sec. 3(i), obtain the expression for $E\left\{V(\mu_j)\right\}$ of the same form as (4.11).

Chapter V

Multiple Character Stratified two phase Sampling Using Conjugate Prior

5.1. Introduction

As in the previous chapter, we consider the two phase sampling in which observations are taken on p characters of the units selected. The population is normally distributed with unknown mean μ_{ij} and unknown variance σ_{ij}^2 for the j^{th} character, $j = 1,2,\ldots,p$. The variation that we consider now is that, instead of non-informative prior on μ_{ij} and σ_{ij}^2, we assume the joint prior distributions of the two parameters, means and variances, are the natural conjugate prior, normal gamma, [Raiffa and Schlaifer (1961) p.300]. We assume also as previously, that the p characters are independently normally distributed. Khan, M.Z.(1976) considered the case of a single character.

5.2. Preposterior Analysis

The first phase observations $x_{ij1}, x_{ij2}, \ldots, x_{ijn_i}$, in the i^{th} stratum, are normally distributed with unknown mean μ_{ij} and unknown variance $\sigma_{ij}^2 = \frac{1}{h_{ij}}$. The likelihood of the first phase observations is the same as (4.3).

However, taking the parameters as (μ_{ij}, h_{ij}) instead of $(\mu_{ij}, \sigma^{2}_{ij})$, (4.3) becomes

$$\prod_{i=1}^{k} (2\pi)^{-\frac{1}{2} n_i} e^{-\frac{1}{2} h_{ij} \sum_{=1}^{n_i} (x_{ij} - \mu_{ij})^2} h_{ij}^{n_i/2} . \quad (5.1)$$

So the kernal is

$$\prod_{i=1}^{k} \exp\left[- \frac{1}{2} h_{ij}(n_i - 1)s^2_{ij} - \frac{1}{2} h_{ij} n_i (\bar{x}_{ij} - \mu_{ij})^2 \right] h_{ij}^{n_i/2}$$

which in the notation of Raiffa and Schlaifer (1961)p.299, becomes

$$\prod_{i=1}^{k} e^{-\frac{1}{2} h_{ij} n_i (\bar{x}_{ij} - \mu_{ij})^2} h_{ij}^{\frac{1}{2}\delta(n_i)} e^{-\frac{1}{2} h_{ij} s^2_{ij} \nu_i^{(1)}} h_{ij}^{k \nu_i^{(1)}/2} \quad \ldots(5.2)$$

where $\nu_i^{(1)} = n_i - 1$, the d.f. of s^2_{ij}, and

$$\delta(n_i^*) = \begin{cases} 0 & \text{if } n_i = o, \text{ the d.f. of } \bar{x}_{ij} \\ 1 & \text{if } n_i > o \end{cases}$$

We assume the joint prior distribution of μ_{ij} and h_{ij} is normal gamma the natural conjugate of (5.2), with parameters m'_{ij}, v'_{ij}, n'_{ij} and ν'_{ij} , where primes means the parameters of the prior distribution, i.e.,

$$\prod_{i=1}^{k} f_{N\gamma}\,(\mu_{ij}, h_{ij} \mid m'_{ij};\ v'_{ij},\ n'_{ij},\ \nu'_{ij})$$

$$= \prod_{i=1}^{k} \left[f_N(\mu_{ij} \mid m'_{ij}, h_{ij} n'_{ij})\, f_{\gamma_2}(h_{ij} \mid v'_{ij},\ \nu'_{ij}) \right]$$

where $f_{N\gamma}$, f_N and $f_{\gamma 2}$ denote the normal gamma, the normal and the gamma-2 distributions, respectively.

The joint posterior distribution of μ_{ij} and h_{ij}, after the first phase observations, is normal gamma and will be proportional to

$$\prod_{i=1}^{k} e^{-\frac{1}{2} h_{ij} n''_{ij} (\mu_{ij} - m''_{ij})}\ h_{ij}^{-\frac{1}{2}\delta(n''_{ij})}\ e^{-\frac{1}{2} \nu''_{ij} v_{ij} h_{ij}}\ h_{ij}^{-\frac{1}{2} \nu''_{ij} - 1} \qquad \ldots (5.3)$$

where $\qquad h_{ij} \geq 0,\ n''_{ij},\ \nu''_{ij} \geq 0. \qquad \ldots (5.4)$

The parameters of the posterior distribution (5.3) are

$$n''_{ij} = n'_{ij} + n_i$$

$$m''_{ij} = \frac{n'_{ij} m'_{ij} + n_i \bar{x}_{ij}}{n'_{ij} + n_i} \qquad \ldots (5.5)$$

$$v''_{ij} = \frac{\left[\nu'_{ij} v'_{ij} + n'_{ij} m'^{2}_{ij}\right] + \left(\nu^{(1)}_{i} s^{2}_{ij} + n_i \bar{x}^{2}_{ij} + n_i \bar{x}^{2}_{ij}\right) - n''_{ij} m''^{2}_{ij}}{\left[\nu'_{ij} + \delta(n'_{ij})\right] + \left[\nu^{(1)}_{i} + \delta(n_i)\right] - \delta(n''_{ij})}$$

...(5.6)

$$\nu''_{ij} = \left[\nu'_{ij} + \delta(n'_{ij})\right] + \left[\nu^{(1)}_{i} + \delta(n_i)\right] - \delta(n''_{ij})$$

It is known that the prior information for the j^{th} character in the i^{th} stratum is equivalent to the information contained in a sample of size n'_{ij} from the population. So it is logical to think that we must have no information if $n'_{ij} = o$, $\nu'_{ij} = o$, and the parameters m''_{ij}, v''_{ij}, n''_{ij} and ν''_{ij} are equal to the statistics $\bar{x}_{ij}$, s^2_{ij}, n_i and $\nu^{(1)}_{i}$ respectively. The above position may be easily verified.

If $y_{ij1}, y_{ij2}, \ldots, y_{ijN_i}$ are the second phase sample observations from the i^{th} stratum for the j^{th} character, the joint distribution of the means $\bar{y}_{1j}, \bar{y}_{2j}, \ldots, \bar{y}_{kj}$ given μ_{ij}, h_{ij} is proportional to

$$\prod_{i=1}^{k} h_{ij}^{\frac{1}{2}N_i} \exp\left[-\frac{1}{2} h_{ij} N_i (\bar{y}_{ij} - \mu_{ij})^2\right] \qquad \ldots(5.7)$$

So the joint distribution of $\bar{y}_{ij}$, μ_{ij}, h_{ij} $(i=1,2,\ldots,k)$ is the product of (5.3) and (5.7).

Then the distribution of the statistic $\bar{y}_{ij}$ will be as shown by Raiffa and Schlaifer (1961), p.304, the general student density, given by

$$p(\bar{y}_{ij} | m''_{ij}, v''_{ij}, n''_{ij}, \nu''_{ij}, N_i, \nu_i^{(2)}) = f_S(\bar{y}_{ij} | m''_{ij}, \frac{n_{u_{ij}}}{v''_{ij}}, \nu''_{ij}) \qquad \ldots(5.8)$$

where $$n_{u_{ij}} = \frac{n''_{ij} \; N_i}{n''_{ij} + N_i} \qquad \ldots(5.9)$$

Thus from (5.8), we have

$$E(\bar{y}_{ij}) = m''_{ij} , \qquad \ldots(5.10)$$

and $$V(\bar{y}_{ij}) = \frac{v''_{ij}}{n_{u_{ij}}} \cdot \frac{\nu''_{ij}}{\nu''_{ij}-2} . \qquad \ldots(5.11)$$

We can see that both $\sum_i \pi_i \mu_{ij}$ posteriori after the first phase and $\sum_i \pi_i \bar{y}_{ij}$ preposteriori before the second phase, have the same expectation $\sum_i \pi_i \; m''_{ij}$, so $\Sigma \; \pi_i \bar{y}_{ij}$ is the preposteriori estimate of $\mu_j = \sum_i \pi_i \mu_{ij}$ and the variance of this estimate is

$$V(\Sigma \; \pi_i \bar{y}_{ij}) = \sum_i \pi_i^2 \; V(\bar{y}_{ij}) = \Sigma \; \pi_i^2 \; \frac{v''_{ij}}{n_{u_{ij}}} \cdot \frac{\nu''_{ij}}{\nu''_{ij}-2}$$

$$= \Sigma\ \pi_i^2 \frac{K_{ij}^2}{n'_{ij}+n_i} \left(1+ \frac{n'_{ij}+n_i}{N_i}\right)$$

$$= \Sigma\ \pi_i^2 \frac{K_{ij}^2}{n'_{ij}+ n_i} + \Sigma\ \pi_i^2\ K_{ij}^2/N_i \qquad \ldots(5.12)$$

where $K_{ij}^2 = \frac{\nu''_{ij}}{\nu''_{ij}-2} \cdot v''_{ij}$.

The first term of (5.12) is a known, constant. Consequently, we are led to the problem of minimizing

$$\Sigma\ \pi_i^2\ K_{ij}^2/N_i \qquad \text{for } j = 1,\ldots,p$$

subject to $\sum_{i=1}^{k} c_i N_i = (1-\beta)c$

and $N_i \geq 0$

which is again in the standard form treated earlier.

5.3. Posterior Analysis.

In the posterior analysis the joint posterior distribution of μ_{ij} and h_{ij} after the first phase observations, viz.(5.3), is taken as the prior distribution before the second phase observation are drawn. Hence all the parameters of (5.3) play the role of the parameters of the prior distribution before the second phase.

Let the second phase observations in the i^{th} stratum be $y_{ij1}, y_{ij2}, \ldots, y_{ijN_i}$. The likelihood is, as in (5.2), proportional to

$$\prod_{i=1}^{k} e^{-\frac{1}{2} h_{ij} N_i (\bar{y}_{ij} - \mu_{ij})^2} \; h_{ij}^{-\frac{1}{2}\delta(N_i)} \; e^{-\frac{1}{2} h_{ij} w_{ij}^2 \nu_i^{(2)}} \; h_{ij}^{\frac{1}{2}\nu_i^{(2)}} \qquad \ldots(5.13)$$

where $\nu_i^{(2)} = N_{i-1}$ = the d.f. of w_{ij}^2, defined earlier.

The joint posterior distribution of μ_{ij}, h_{ij} is again normal gamma obtained by multiplying (5.13) and (5.3), with parameters

$$n_{ij}''' = n_{ij}'' + N_i$$

$$m_{ij}''' = \frac{n_{ij}'' m_{ij}'' + N_i \bar{y}_{ij}}{n_{ij}'' + N_i}$$

$$v_{ij}''' = \frac{[\nu_{ij}'' v_{ij}'' + n_{ij}'' m_{ij}''^2] + [\nu_i^{(2)} w_{ij}^2 + N_i \bar{y}_{ij}^2] - n_{ij}''' m_{ij}'''^2}{[\nu_{ij}'' + \delta(n_{ij}'')] + [\nu_i^{(2)} + \delta(N_i)] - \delta(n_{ij}''')} \qquad \ldots(5.14)$$

$$\nu_{ij}''' = [\nu_{ij}'' + \delta(n_{ij}'') + \nu_i^{(2)} + \delta(N_i)] - \delta(n_{ij}''') .$$

The marginal distribution of μ_{ij}, obtained by integrating out the nuisance parameter h_{ij}, is the Student's distribution

$$p(\mu_{ij}|m_{ij}''',v_{ij}''',n_{ij}''',\nu_{ij}''')= f_S(\mu_{ij}|m_{ij}''',\frac{n_{ij}'''}{v_{ij}'''}, \nu_{ij}''').$$

Hence we have

$$E(\mu_{ij}|m_{ij}''',v_{ij}''',n_{ij}''',\nu_{ij}''') = m_{ij}''',$$

$$V(\mu_{ij}|m_{ij}''',v_{ij}''',n_{ij}''',\nu_{ij}''')=\frac{\nu_{ij}'''}{\nu_{ij}'''-2} \cdot \frac{v_{ij}'''}{n_{ij}'''} . \qquad \ldots(5.14)$$

Since v_{ij}''' involves the second phase observations (5.14) can not be used for the second phase allocation. So we replace v_{ij}''' by its expectation in the posterior variance of μ_{ij} in (5.14). As the marginal distribution of v_{ij}''' is inverted β eta, using the result given by Raiffa and Schlaifer, we have

$$E(v_{ij}''') = \frac{\nu_{ij}'' v_{ij}''}{\nu_{ij}'''} \cdot \frac{\frac{1}{2}\nu_{ij}'''-1}{\frac{1}{2}\nu_{ij}''-1} .$$

This gives the expected $V(\mu_{ij})$ (the expectation is taken over the future distribution of Y) after simplification as

$$E\,|V(\mu_{ij}\,|m_{ij}''',v_{ij}''',n_{ij}''',\nu_{ij}''')| = \frac{\nu_{ij}''v_{ij}''}{\nu_i'''-2} \cdot \frac{\frac{1}{2}\nu_{ij}'''-1}{\frac{1}{2}\nu_{ij}''-1} \cdot \frac{1}{n_{ij}'''}$$

$$= K_{ij}^2 \frac{1}{n_{ij}'''} .$$

So the expected posterior variance of $\mu_j = \Sigma\, \pi_i \mu_{ij}$ is

$$= \sum_i \pi_i^2 \quad V(\mu_{ij})$$

$$= \Sigma\, \pi_i^2\, K_{ij}^2 \quad \frac{1}{n_{ij}' + n_i + N_i}$$

$$= \sum_{i=1}^{k} \frac{U_{ij}}{n_{ij}'+n_i+N_i} \text{ where } U_{ij} = \pi_i^2 K_{ij}^2 .$$

By putting $n_{ij}' + n_i + N_i = x_i$, the allocation problem again reduces to our standard form given in (2.1) to (2.3).

R E F E R E N C E S

1. Ahsan, M.J. and Khan, S.U. (1977). '' Optimum allocation in Multivariate Stratified random sampling using prior information''. J.I.S.A. Vol.15 pp (57-67).

2. Benayonn, R, J.de Montcolfier and J.Tergny (1971). "Linear Programming with multiple objective function, STEP method (STEM) ''.

3. Benayonn, R et J.Tergny (1969). '' Criters Multiple E_n Programmation Mathematique Une Solution dans le cas Lineare''. R.I.R.O. 2, 31-56.

4. Box, G.E.P. and Tiao, G.C (1973). '' Bayesian Inference in Statistical Analysis''. Addison-Wesley.

5. Cochran, W.G.(1977). '' Sampling Technique ''. John Wiley and Sons.

6. Collatz, L. and W.Wetterling (1975).'' Optimization Problems ''. Springer-Verlag.

7. Dakin, R. (1965). '' A Tree search Algorithm for Mixed Integer Programming Problem''. Computer Journal 8, 250-255.

8. Draper, N.R. and Guttman, Irwin (1968).'' Some Bayesian Stratified two-phase sampling results''. Biometrika vol.55 p.131-139.

9. Ericson, W.A.(1965). '' Optimum Stratified Sampling Using prior information ''. J.A.S.A. vol.60, pp750-772.

10. Khan, M. Z.(1976). '' Optimum Allocation in Bayesian Stratified Two phase Sampling. J.I.S.A. vol.14 pp. 65-74.

11. Krabs, W. (1979). '' Optimization and Approximation''. John Wiley.

12. Land, A.H. and A.G.Doig (1960).'' An Automatic method for solving Discrete Programming Problems''. Econometrica, 28 , pp. 497 - 520.

13. Raiffa, H. and Schlaifer (1961). '' Applied Statistical Decision Theory''. Boston,Division of Research, Harward Business School.

14. Rao, C. R. (1973). '' Linear Statistical Inference and its Applications''. Wiley.

15. Roy, B. (1971). '' Problems and Methods with Multiple Objective functions ''. Mathematical Programming 1,239-366.
